



Early Journal Content on JSTOR, Free to Anyone in the World

This article is one of nearly 500,000 scholarly works digitized and made freely available to everyone in the world by JSTOR.

Known as the Early Journal Content, this set of works include research articles, news, letters, and other writings published in more than 200 of the oldest leading academic journals. The works date from the mid-seventeenth to the early twentieth centuries.

We encourage people to read and share the Early Journal Content openly and to tell others that this resource exists. People may post this content online or redistribute in any way for non-commercial purposes.

Read more about Early Journal Content at <http://about.jstor.org/participate-jstor/individuals/early-journal-content>.

JSTOR is a digital library of academic journals, books, and primary source objects. JSTOR helps people discover, use, and build upon a wide range of content through a powerful research and teaching platform, and preserves this content for future generations. JSTOR is part of ITHAKA, a not-for-profit organization that also includes Ithaka S+R and Portico. For more information about JSTOR, please contact support@jstor.org.

SUMMARY AND GROUP CORRELATION

The possibility of finding the arithmetic mean of several groups separately, and combining the data in order to find the arithmetic mean of the combined groups, is well known and is frequently used to save the labor of additional computations. It may not be generally known that an analagous process is available in the case of the coefficient of correlation.

Let x and y be the deviations, and A and B the averages of two series, the original values being expressed by X and Y .

$$\begin{aligned}\text{Then} \quad \Sigma xy &= \Sigma(X-A)(Y-B) \\ &= \Sigma XY + nAB - A\Sigma Y - B\Sigma X \\ &= \Sigma XY + \frac{n(\Sigma X)(\Sigma Y)}{n} - \frac{(\Sigma X)(\Sigma Y)}{n} - \frac{(\Sigma X)(\Sigma Y)}{n} \\ &= \Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n} \\ \Sigma x^2 &= \Sigma(X-A)^2 \\ &= \Sigma X^2 - 2A\Sigma X + nA^2 \\ &= \Sigma X^2 - 2\frac{(\Sigma X)(\Sigma X)}{n} + n\frac{(\Sigma X)^2}{n^2} \\ &= \Sigma X^2 - \frac{(\Sigma X)^2}{n}\end{aligned}$$

$$\text{and, similarly, } \Sigma y^2 = \Sigma Y^2 - \frac{(\Sigma Y)^2}{n}.$$

$$\begin{aligned}\text{Hence} \quad r &= \frac{\Sigma xy}{n\sigma_x\sigma_y} = \frac{\Sigma xy}{\sqrt{\Sigma x^2\Sigma y^2}} \\ &= \frac{\Sigma XY - \frac{(\Sigma X)(\Sigma Y)}{n}}{\sqrt{\left[\Sigma X^2 - \frac{(\Sigma X)^2}{n}\right]\left[\Sigma Y^2 - \frac{(\Sigma Y)^2}{n}\right]}} \\ &= \frac{n\Sigma XY - (\Sigma X)(\Sigma Y)}{\sqrt{[n\Sigma X^2 - (\Sigma X)^2][n\Sigma Y^2 - (\Sigma Y)^2]}}.\end{aligned}$$

Since X and Y may be any scalar values, the items may be grouped into classes, and arbitrary values may be assigned to the classes, *e. g.*, $-2, -1, 0, +1, +2, \dots$

If in two or more groups the same arbitrary class values are used, the number of items in one group being n_1 , in another n_2 , etc., the correlation coefficient for each group may be obtained separately, and by a simple computation the correlation coefficient for the combined groups may be obtained:

$$r = \frac{(n_1 + n_2 + \dots)\Sigma(\Sigma XY) - [(\Sigma X)(\Sigma Y)]}{\sqrt{\{(n_1 + n_2 + \dots)\Sigma(\Sigma X^2) - \Sigma[(\Sigma X)^2]\} \{(n_1 + n_2 + \dots)\Sigma(\Sigma Y^2) - \Sigma[(\Sigma Y)^2]\}}}.$$

HERBERT A. STURGES

University of Washington